Low-Cost Security of IoT Sensor Nodes with Rakeness-Based Compressed Sensing: Statistical and Known-Plaintext Attacks

Low-cost Security of IoT Sensor Nodes With Rakeness-Based Compressed Sensing: Statistical and Known-Plaintext Attacks

Mauro Mangia, *Member, IEEE*, Fabio Pareschi, *Member, IEEE*, Riccardo Rovatti, *Fellow, IEEE*, Gianluca Setti, *Fellow, IEEE*

***Abstract*—Compressed Sensing has been proposed to both yield low-cost compression and low-cost encryption. This can be very useful in the design of sensor nodes with a limited resource budget whose acquisition must be kept as private as possible. We here analyze the susceptibility of Compressed Sensing stages that are optimized to maximize compression performance by rakeness-based design to ciphertext-only and known-plaintext attacks. A trade-off between compression and security is highlighted. Notwithstanding such a trade-off, rakeness-based Compressed Sensing exhibits a noteworthy robustness to classical attacks.**

I. Introduction

THE TERM *Internet of Things* (IoT) first appeared in 1999 and rapidly evolved into the label for the collective systems arising from the exchange of information (mediated by a communication network) between devices whose purpose is to interact with the physical world in most diverse ways.

The aim of these systems is to provide *smart* functionalities that would not be possible without the acquisition and exchange of such information and, today, we begin to design and deploy smart grids, smart homes, smart water networks, smart transportation infrastructures, etc., that deeply rely on the intertwining of diverse physical interactions and information processing.

A key ingredient in IoT aggregates is their ability of autonomously gathering information about the real world by means of a potentially large number of sensors either embedded in objects or deployed ad-hoc. These sensors typically work on extremely low resource budgets in terms of available power, weight, etc. To limit their invasiveness with respect to the phenomena they observe, they communicate the acquired data wirelessly to the other elements of the IoT ensemble.

Hence, ability of designing extremely parsimonious wireless sensing *nodes* is fundamental for the development of IoT systems. For example, low-level processing and elementary signal compression is sought to limit the data rate of the transmission that is often the main cause of power consumption.

F. Pareschi and G. Setti are with the Department of Engineering, University of Ferrara, 44122 Ferrara, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: fabio.pareschi@unife.it; gianluca.setti@unife.it).

M. Mangia and R. Rovatti are with the Department of Electrical, Electronic, and Information Engineering, University of Bologna, 40136 Bologna, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: mauro.mangia2@unibo.it; riccardo.rovatti@unibo.it).

Yet, such a need for parsimony must cope with an intrinsic downside of every distributed system in which information is transferred between its components through public channels: a *privacy issue*. Malicious entities may be interested in interfering with the data flow that makes IoT work, with the aim of capturing sensible information (e.g., biological data of patients being monitored), predicting system behavior (e.g., knowing in advance whether or not an alarm will be triggered as a consequence of some event), or even altering it (e.g., preventing automatic responses to threats). In principle, *every sensor node should accommodate an encryption stage designed to address the trade-off between the level of security that is needed and the complexity of the stage itself.*

To address the simultaneous need for compression and privacy, the use of *Compressed Sensing* (CS) has been recently proposed as a way to both reduce the amount of data to transmit [1], [2] and make them computationally secure.

The main advantage of the approach is that CS entails only very simple processing. The downside is that security is not perfect and has to be assessed in terms of the computational power needed by an attacker to break the encryption and reveal the data. Such an analysis has been carried out for the classical CS setting [3][4], but it is still an open problem for the improved version of CS proposed in [5][6], in which compression performance is significantly increased by adapting to second-order features of the signal to acquire.

This paper fills the gap by analyzing the effect of such an optimization on the security offered by the CS stage. In particular we consider Ciphertext Only Attacks (COAs) in which an eavesdropper observes only encoded signals and tries to infer by statistical means some of the features of the true signal. We also tackle Known Plaintext Attacks (KPAs) in which the attacker has access to both the original and encoded version of the signal in certain time windows, and tries to break the encryption of future waveforms. This is an attack that is particularly important for sensing nodes since the attacker may think of temporary deploying an identical sensor close to the one to attack to obtain the unencoded reading. These scenarios are depicted in Figure 1.

Results shows that, especially for KPAs, there is a trade-off between the compression performance and obtained security. Yet, even when compression is significantly increased with respect to classical CS, the computational resources needed by an attacker to reveal the original signal appear to be well beyond what can be reasonably spent to obtain protected

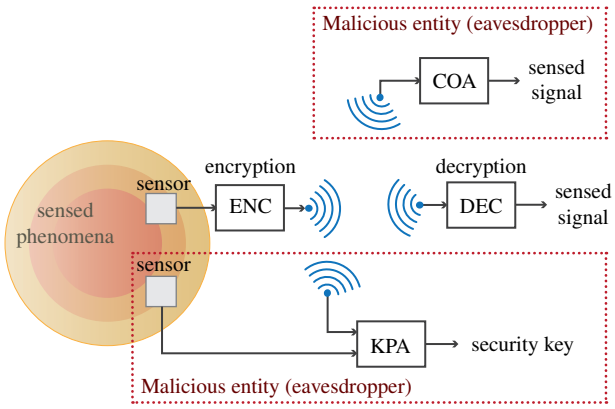

Fig. 1. Block scheme of a WSN where the encoder (ENC) performs the encryption and the decoder (DEC) retrieves the acquired signal. The picture also shows two malicious users, one trying to recover the security key by a know plaintext attack (KPA) (acquiring the same signal with another sensor and eavesdropping the encrypted data) and a second one trying to recover the sensed signal only by eavesdropping the encrypted data (the ciphertext only attack, COA).

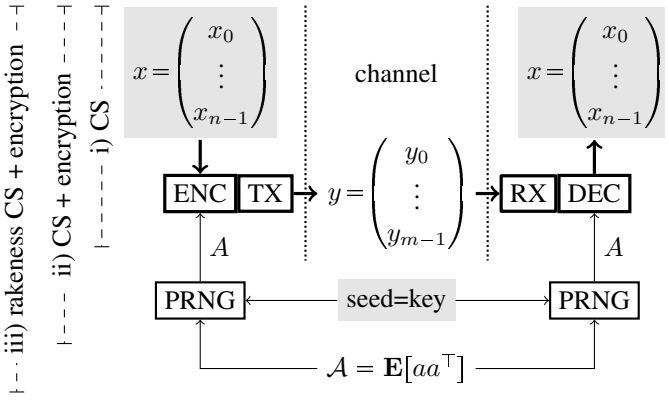

Fig. 2. A scheme of the possible uses of CS in a sensor node: i) pure CS encodes a vector of n samples into m scalars that can be transmitted and used to reconstruct the original signal; ii) if the seed of the PRNG used to generate the encoding matrix is kept secret, CS is also equivalent to a private-key encryption: iii) if compression is increased by means of rakeness based CS the matrix $\mathcal{A}$ is an additional information on the encoding that is publicly available.

information with limited value like most sensor acquisition are.

The paper is organized as follows. Section II recalls the fundamental concepts of CS and describes the interplay between CS and encryption. Section III summarizes the so-called rakeness-based design flow that optimizes compression performance of CS by adapting to the second-order features of the signal to acquire. Section IV describes the model we use to perform the cryptanalysis. Section V applies such a model to COAs while Section VI applies that to KPAs.

Some numerical examples are detailed in Section VII, in which we consider both a class of synthetic signals allowing the description of the compression-security trade off, and real-world Electro Cardio Graphic (ECG) signals. In both cases the effectiveness of the theoretical guarantee is assessed and some estimation of the computational effort needed to break the encryption is given.

Based on all the above, some conclusion are finally drawn.

The Appendix contains the proofs of few Lemmas that are needed to develop the theoretical part.

II. Compressed Sensing and Encryption

The input waveform is represented by a set of n samples collected in a *signal* $x = (x_0, \dots, x_{n-1})^\top \in \mathbb{R}^n$. The acquisition of such a waveform can benefit from CS when x is known to be κ-*sparse*, i.e., when there is an n-dimensional *sparsity basis* $S \in \mathbb{R}^{n \times n}$ such that expressing $x = Ss$ yields a vector $s \in \mathbb{R}^n$ with not more than $\kappa \ll n$ non-zero components.

The number of true degrees of freedom in x is therefore considerably smaller than n. Leveraging this property, fundamental results [7] have shown that the signal can be captured by a set of $m < n$ properly designed linear *measurements*. We arrange these measurements in the m-dimensional vector $y = (y_0, \dots, y_{m-1})^\top \in \mathbb{R}^m$, obtained by applying a *projection matrix* $A \in \mathbb{R}^{m \times n}$ to x, i.e., $y = Ax = ASs$. Passing from x to y can be seen as an *encoding* of the signal to acquire into m scalar quantities.

CS theory [7], [8] guarantees that s (and thus x) can be recovered from y despite the fact that A (and thus AS) yields a dimensionality reduction, provided that $m = \mathcal{O}(\kappa \log n)$.

This is done by exploiting the *a priori* knowledge on the sparsity of s to identify it among all the vectors satisfying of $y = ASs$. Algorithms performing this *sparse signal recovery* have been explored and improved in recent years [7], [9], [10], [11] and can be seen as a *decoding* stage inverting the linear encoding.

This said, it is fair to state that most of the practical interest in CS comes from two key facts. First, although theoretical upper bounds exist on the error committed by signal recovery algorithms depending on the features of A and S, and on the amount of noise, their actual performance largely exceeds what is predicted by formal guarantees, yielding signal recovery from a much smaller number of measurements [6]. Second, the mathematical conditions that allow sparse signal recovery can be matched (with *very high* probability) by simply drawing A at random. Although theoretical guarantees depend on the choice of specific distributions [12], in practice a wide class of random matrices allows for effective signal recovery [8].

These two points, jointly with the intrinsic simplicity of the processing that CS requires at the sensing side (the computation of Ax), suggests that CS can be a valid option to reduce resource needs in sensing nodes since it provides a non-negligible compression at a very low cost.

The general scheme is the one in the upper part of Figure 2, i.e., that corresponding to the side label "i) CS". The signal vector is encoded and transmitted. Since encoding provides compression, communication resources like time and energy are reduced [13], [14]. Thanks to CS, this comes at the expense of an extremely light computation that may be also optimized [15] to become almost negligible, starting from the common assumption $A \in \{-1, 0, +1\}^{m \times n}$ that reduces multiply-and-accumulate units to simple signed sums. In the following we will assume $A \in \{-1, +1\}^{m \times n}$.

Further to this, researchers have realized that CS can be exploited to provide also some form of security for the

acquired data. In fact, recovery needs the knowledge of A and this suggests looking at the encoding process as a private-key encryption stage for which x is the plaintext, y is the ciphertext and A is the shared secret. This is the core idea in [17][18][19][20][21].

Practical implementations may follow the scheme at the top of Figure 2 in the part labelled "ii) CS+encryption" in which the actual *key* is the seed of a Pseudo-Random Number Generator (PRNG) producing A (a different instance for each set of n samples) both at the encoder and at the decoder. Note that, according to the Kerchoff's principle, the only secret quantities in such a scheme are those with a shaded background, i.e., the key and, consequently, the signal at the sensing node and at the receiving hub. On the contrary measurements, encoding strategy and PRNG structure are known to attackers.

The robustness of such an encryption to some classical attacks has been investigated (see, for example, [22], [20] and [3]) for systems in which the PRNGs expand the key in a stream of statistically independent symbols ± 1. Ciphertext-only attacks (COAs) have been shown to be ineffective since, when n is large, they may only reveal the average energy of x. Hence, CS-based encryption enjoys *asymptotic circular secrecy*, i.e., it is asymptotically Shannon-secure [20] when leaking the energy of the ciphertext is not an issue.

Known-plaintext attacks (KPAs) have been also considered, in which the attacker knows both x and y at certain instants in time and aims at retrieving the corresponding A that is the output of the PRNG. From such an output the attacker hopes to identify the key and be able to seed a copy of the PRNG to anticipate future encoding matrices and break encryption. In conventional schemes, security hinges on the the proper design of the PRNG. Yet, here we cannot exploit any known techniques to deploy cryptographically-secure PRNG since i- we aim at outputting non-independent symbols and thus need specialized generation schemes, ii- coherently with the low-cost sensor node setting typical of IoT, the resources devoted to key expansion must be extremely limited. In this case, robustness comes from the fact that each plaintext-ciphertext is compatible with an enormous number of antipodal matrices among which the true one sits like an indistinguishable *straw* in a haystack. Hence, even if going from the PRNG output to the key is not and hard task, the KPA still fails since inferring the PRNG output from the plaintext-ciphertext pair may be practically unfeasible.

III. Rakeness-based CS in a nutshell

Rakeness-based design of CS improves on the classical approach and on all known approaches to CS optimization by noting that real-world signals x are usually not only sparse, but also non-white or *localized* [6].

Their energy is anisotropically distributed in the signal space, or, more formally, assuming $\mathbf{E}[x] = 0$, their correlation matrix $\mathcal{X} = \mathbf{E}[xx^\top]$ is not a multiple of the identity (from now on, calligraphic typefaces indicate correlation matrices). The degree of such a localization can be assessed by computing how much the eigenvalues η_j of $\mathcal{X}$ deviate from the isotropic case, i.e.,

$$\mathfrak{L}_x = \sum_{j=0}^{n-1} \left(\frac{\eta_j}{\sum_{k=0}^{n-1} \eta_k} - \frac{1}{n} \right)^2 = \frac{\mathrm{tr}\left(\mathcal{X}^2\right)}{\mathrm{tr}^2(\mathcal{X})} - \frac{1}{n} \qquad (1)$$

where $\mathrm{tr}(\cdot)$ stands for matrix trace and one can show that $0 \leqslant \mathfrak{L}_x \leqslant 1 - \frac{1}{n}$.

When $\mathfrak{L}_x = 0$ all the eigenvalues of $\mathcal{X}$ are equal and the signal is white. On the contrary, when $\mathfrak{L}_x > 0$ the larger eigenvalues of $\mathcal{X}$ correspond to directions along which the signal is more likely to put energy. These are directions along which projections may want to focus, though they should remain able to explore other directions that on the average are less energetic but may be important to reconstruct individual instances of x.

This is what rakeness-based design of A does [5], [6], assuming that A is made of independent and identically distributed rows, whose entries may be given a non-white statistics to improve acquisition performance. Formally speaking, indicate with $a = (a_0, \ldots, a_{n-1})^\top$ the random column vector corresponding to a generic row of A so that $a^\top x$ is the generic measurement.

Then define the *rakeness* of a with respect to x as $\rho(a, x) = \mathbf{E}_{a,x}\left[\left(a^\top x\right)^2\right]$, i.e., as the average of the energy that measurements capture from the signal. If $\mathcal{A} = \mathbf{E}_a[aa^\top]$, then $\rho(a, x) = \mathrm{tr}\left(\mathcal{A}\mathcal{X}\right)$.

The correlation matrix $\mathcal{A}$ of a also controls its localization since

$$\mathfrak{L}_a = \frac{\mathrm{tr}(\mathcal{A}^2)}{\mathrm{tr}^2(\mathcal{A})} - \frac{1}{n}. \qquad (2)$$

With this, the idea of favoring most energetic directions while not neglecting the possibility of spanning the whole signal space can be translated in finding the correlation matrix $\mathcal{A}$ that maximizes $\rho(a, x)$ when $\mathfrak{L}_a$ does not exceed a certain threshold. To parameterize the localization of the projections with that of the signal we required $\mathfrak{L}_a \leqslant \ell \mathfrak{L}_x$ for some $0 \leqslant \ell \leqslant 1$.

Note that setting $\ell = 0$ implies $\mathfrak{L}_a = 0$ so that the sensing vectors are forced to be white and, since they are antipodal, made of independent entries as in classical CS.

The rakeness maximization problem with localization constraint can be solved analytically [5]. If $\mathcal{X}$ is spectrally decomposed as $\mathcal{X} = \sum_{j=0}^{n-1} \eta_j q_j q_j^\top$ with η_j the eigenvalues sorted in non-increasing order and q_j the corresponding orthonormal eigenvectors, then the optimized correlation matrix is $\mathcal{A} = \sum_{j=0}^{n-1} \lambda_j q_j q_j^\top$ for certain eigenvalues λ_j that, in the simplest case, are given by

$$\lambda_j = \left(1 - \sqrt{\ell}\right) + n \frac{\eta_j}{\sum_{k=0}^{n-1} \eta_k} \sqrt{\ell} \qquad (3)$$

where values of ℓ around $1/4$ are usually employed. The resulting maximal rakeness is

$$\rho^*(a, x) = \sum_{j=0}^{n-1} \eta_j \lambda_j = \left(1 + n\sqrt{\ell}\mathfrak{L}_x\right) \sum_{j=0}^{n-1} \eta_j \qquad (4)$$

Once that $\mathcal{A}$ is known, a number of methods exist to generate antipodal vectors a featuring a correlation matrix

as close as possible to $\mathcal{A}$ [23], [24], [25]. The practical effect of using the resulting vectors as rows of A is that the amount of information that each measurement carries about the signal increases. Hence, less measurements are needed to allow signal reconstruction and compression is sensibly increased with respect to conventional purely random approaches and also with respect to other general purpose adaptation methods available in the Literature [26].

There is, however, a drawback which is of particular importance for the analysis in this paper: it is no longer true that the entries of A are independent since its rows are characterized by a correlation matrix $\mathcal{A}$. This situation is represented in Figure 2 globally labelled "iii) rakeness CS + encryption" where we highlight that the $\mathcal{A}$ controlling the PRNGs both at the encoder and at the decoder is known to attackers.

Intuitively, more information for the attackers means less security, and the aim of the following discussion is to quantify such a loss and ascertain that it does not impair the usability of CS-based encryption in low-cost sensor nodes.

IV. Model and assumptions for cryptanalysis

Most of our considerations will be *for large* n, i.e., in the asymptotic condition $n \to \infty$ that requires increasingly long signals x and increasingly long rows a.

To cope with this we assume that the samples x_j come from a zero-mean, exponentially mixing (and thus ergodic and stationary) stochastic process with power spectrum $S_x(f)$ and power

$$\mathbf{E}[x_j^2] = \int_{-1/2}^{1/2} S_x(f)\mathrm{d}f = W_x$$

We know from the main Theorem in [27, chapter 5] that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \eta_j = \int_{-1/2}^{1/2} S_x(f)\mathrm{d}f = W_x \tag{5}$$

As a regularity assumption we additionally require that $S_x(f)$ is square summable so that from the same Theorem

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \eta_j^2 = \int_{-1/2}^{1/2} S_x^2(f)\mathrm{d}f = \xi_x W_x^2 \tag{6}$$

for some $\xi_x < \infty$. Non-white signals have $\xi_x > 1$.

Since the x_j model samples from real world-quantities, it is also very sensible to assume

$$\mathbf{E}\left[\,|x_j|^{4+p}\right] < \infty \tag{7}$$

for some $p > 0$. Note that for random variables x_j with a bounded support (7) holds for any $p > 0$.

As far as rows of A are concerned, we assume that a is made of entries coming from a process that is also zero-mean and exponentially mixing with power spectrum $S_a(f)$. Since a is antipodal we know that all the diagonal entries of $\mathcal{A}$ are equal to 1 and thus

$$\frac{1}{n}\sum_{j=0}^{n-1} \lambda_j = 1$$

Yet, a straightforward calculation exploiting (6) and (3) yields

$$\xi_a = \int_{-1/2}^{1/2} S_a^2(f)\mathrm{d}f = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \lambda_j^2 = 1 + (\xi_x - 1)\,\ell \tag{8}$$

It is worthwhile stressing that (6) and (8) are regularity conditions in that, for example, exclude delta-like pulses in the power spectra. Moreover, since both the process generating x and the one generating a are assumed to be stationary, their autocorrelation functions $C_x(\tau) = \mathbf{E}[x_0 x_\tau]$ and $C_a(\tau) = \mathbf{E}[a_0 a_\tau]$, whose values populate the Töplitz matrices $\mathcal{X}$ and $\mathcal{A}$, are the Fourier inverses of the corresponding spectra. For these functions, (6) and (8) jointly with the Parseval equality give

$$\begin{aligned}\sum_{\tau=-\infty}^{\infty} C_x^2(\tau) &= \xi_x W_x^2 < \infty \\ \sum_{\tau=-\infty}^{\infty} C_a^2(\tau) &= 1 + (\xi_x - 1)\,\ell < \infty\end{aligned}$$

that further confirm that we are dealing with weakly correlated signals.

When needed, in the following we may work with vectors that are normalized to have unit power, i.e. with $\bar{x} = x/\sqrt{nW_x}$, $\bar{a} = a/\sqrt{n}$, and $\bar{y} = \bar{a}^\top \bar{x}$.

In the next Sections we see that the parameters ℓ and $\mathfrak{L}_x$, W_x and ξ_x control the statistics of the ciphertext and thus its susceptibility to COAs, as well as the probability of success of a KPA.

V. Ciphertext statistics and COAs

Robustness againt COAs depends on the statistics of the measurements that, in their normalized form, obey the Central Limit Theorem for normalized sums of *non-independent* variables. To show this we will mainly use results from [28].

Consider any two measurements $y' = a'^\top x$ and $y'' = a''^\top x$. Since $\mathbf{E}[a'] = \mathbf{E}[a''] = 0$ both measurements have zero mean. Thanks to the independence of a', a'' and x, the covariance between measurements is $\mathbf{E}[y'y''] = \mathbf{E}[a'^\top x x^\top a''] = \mathbf{E}[a']^\top \mathbf{E}[xx^\top]\mathbf{E}[a''] = 0$.

Moreover, choose any two coefficients z' and z'', build the random variable $Y = z'y' + z''y'' = (z'a' + z''a'')^\top x$ and normalize it in $\bar{Y} = Y/\sqrt{n}$ such that

$$\bar{Y} = \frac{1}{\sqrt{n}}\sum_{j=0}^{n-1} (z'a'_j + z''a''_j)x_j = \frac{1}{\sqrt{n}}\sum_{j=0}^{n-1} X_j \tag{9}$$

where the composite random variables $X_j = (z'a'_j + z''a''_j)\bar{x}_j$ remain implicitly defined. Clearly, $\mathbf{E}[X_j] = 0$ and from our assumption (7) we get $\mathbf{E}[|X_j|^{4+p}] < \infty$. Moreover, if the process generating a' and a'' and the one generating x are exponentially mixing, so is the process generating the X_j. Hence, our normalized sum (9) satisfies all the assumptions of [28, Theorem 4] and, as $n \to \infty$, $\bar{Y}$ asymptotically behaves like a Gaussian random variable.

Since the two coefficients are arbitrary, the asymptotic Gaussianity of $\bar{Y}$ implies that $y'/\sqrt{n}$ and $y''/\sqrt{n}$ are asymptotically jointly Gaussian. Since y' and y'' have zero covariance they are also independent and are equally distributed as Gaussians with a variance that is the average energy of the normalized measurements $\mathbf{E}[{y'}^2]/n = \mathbf{E}[{y''}^2]/n$. Since the average energy of measurements due to rakeness-based design is given by (4) we have from (6) that

$$\frac{\mathbf{E}[{y'}^2]}{n} = \frac{\mathbf{E}[{y''}^2]}{n} = \left(1 + n\sqrt{l}\mathfrak{L}_x\right)\frac{1}{n}\sum_{j=0}^{n-1}\eta_j$$

whose limit can be computed thanks to (1), (5), and (6) to yield

$$\frac{y'}{\sqrt{n}} \overset{n\to\infty}{\sim} \frac{y''}{\sqrt{n}} \overset{n\to\infty}{\sim} \mathcal{N}\left(0, \left(1 + \sqrt{\ell}\left(\xi_x - 1\right)\right) W_x\right) \quad (10)$$

The same [28, Theorem 4] tells us that convergence speed is only slightly impaired by the dependency between the coefficients of a due to rakeness-based design. In fact, the difference between the cumulative distribution function of the measurements $\Phi_{y/\sqrt{n}}$ and the cumulative distribution function of the limit Gaussian $\Phi_{0,(1+\sqrt{\ell}(\xi_x-1))W_x}$ is bounded by

$$\left\|\Phi_{y/\sqrt{n}} - \Phi_{0,(1+\sqrt{\ell}(\xi_x-1))W_x}\right\|_\infty \leqslant \Theta\frac{\log n}{\sqrt{n}}$$

for some constant Θ decreasing when p increases, that may be compared with the $\mathcal{O}(1/\sqrt{n})$ trend of the classical Berry-Esseen bound [29].

This shows that rakenes-based CS used as an encryption stage is asymptotically circularly secret as defined in [20], since $\left(1 + \sqrt{l}\left(\xi_x - 1\right)\right) W_x$ is the only information that an attacker may infer from the observation of the ciphertexts y.

The results in [20] are a special case of (10). In fact, recall that an A made of i.i.d. antipodal entries is obtained by setting $\ell = 0$ and in that case, the information that leaks is W_x.

What we get from (10) is that asymptotic circular security holds also adopting a rakeness-based design flow as described in Section III.

Note also that, since the measurements are asymptotically independent and identically distributed, considering the statistic of more than one measurement does not give any additional information on the hidden vector x.

VI. Success probability of KPAs

In this case the attacker knows a plaintext-ciphertext pair $(x, y = Ax)$ and aims at reconstructing A. Despite it is true that the very same matrix will not be used for future encodings, its knowledge is still of use for an attacker. In fact, subsequent matrices are generated by the same PRNG that can be attacked to retrieve its seed and thus the key. If this happens all subsequent ciphertexts can be decoded. The PRNG has to presumably be a simple one, given the limited resources available at the sensing node, but for the rakeness approach it must be capable to generate sequences with a prescribed second order statistics. A good candidate is described in [23].

The reconstruction of A must proceed row by row since, by assumption, rows are generated as independent realizations of the same process with correlation matrix $\mathcal{A}$. Hence, the attack concretizes in solving m equations of the kind $y = a^\top x$ when the scalar measure y and the signal vector x are given (note that from now on, since we concentrate on only one measure, y is a scalar).

In principle, solving such an equation is not a difficult task. Yet, the solution is not unique and, among the extremely huge amount of solutions, the chance of hitting one that is at least close to the true one is negligible.

In [3], the authors show that if A and thus a is made of antipodal i.i.d. entries, the number of solutions is large enough to deter an attacker as no side-information can be exploited to prioritize the search for the true one.

Yet, if $\mathcal{A}$ is not the identity, the attacker may exploit such an information.

The strongest possible attack is able to generate guesses a_{guess} that simultaneously satisfy all the available information, i.e., they satisfy the measurement equation $y = a_{\text{guess}}^\top x$ and their second-order statistics is regulated by the known correlation matrix $\mathcal{A}$.

Since row guessing must be repeated for every row of A, the attacker benefits also from the fact that we adopt rakeness-based CS to increase the compression rate $\texttt{CR} = n/m$, i.e., to decrease m while not impairing reconstruction performance. With rakeness-based CS, CR is larger and thus the number of rows of A to guess is smaller with respect to classical CS, possibly easing the attacker task proportionally with the increase in CR.

Our theoretical development focuses on the probability that the informed guesses produced by the attacker are close to the true row.

From this point of view, the chances of success of a KPA to each row depend on the probability that two sensing vectors a (the one used by the encoder) and a_{guess} (the one guessed by the attacker) independentely generated by the same antipodal process characterized by the correlation matrix $\mathcal{A}$, are very similar. For antipodal vectors, the most natural measure of similarity is their Hamming distance $\Delta(a, a_{\text{guess}})$, i.e.,the number of entries in which a and a_{guess} disagree.

Regrettably the geometry of $\Delta(a, a_{\text{guess}})$ is not trivial and exploiting the intuition in [32] we will derive approximate bounds on the probability of a KPA success noting that

$$\frac{1}{n}\Delta(a, a_{\text{guess}}) = \frac{1}{4}\left\|\frac{a}{\sqrt{n}} - \frac{a_{\text{guess}}}{\sqrt{n}}\right\|^2 = \frac{1}{4}\left\|\bar{a} - \bar{a}_{\text{guess}}\right\|^2 \quad (11)$$

and relaxing the antipodality constraint on the normalized sensing vectors to allow some analytical considerations.

The geometrical structure of the KPA attack we analyze is reported, in normalized form, in Figure 3-(a). The vector $\bar{x}$, the vector used by the encoder ($\bar{a}'$) and the one guessed by the attacker ($\bar{a}''$) are such that $(\bar{a}')^\top\bar{x} = (\bar{a}'')^\top\bar{x}$, i.e. they form the same angle with $\bar{x}$. Since $\|\bar{a}'\| = \|\bar{a}''\| = 1$ both $\bar{a}'$ and $\bar{a}''$ lie on the intersection between the unit n-dimensional sphere and a cone centered on $\bar{x}$, i.e. on a $n-1$ dimensional sphere orthogonal to $\bar{x}$, centered in $c = \bar{y}\bar{x}/\|\bar{x}\|^2$, and with

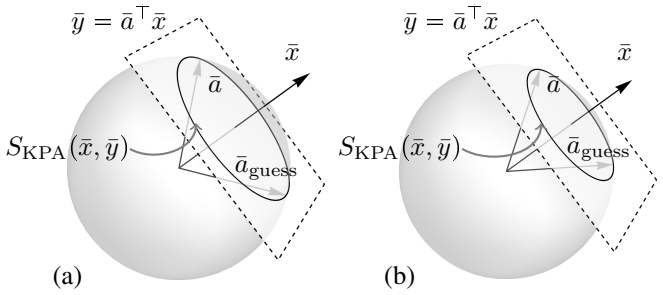

Fig. 3. The first effect of rakeness-based design on KPA susceptibility: the angle between x and the sensing vectors is typically smaller so that guesses are typically closer.

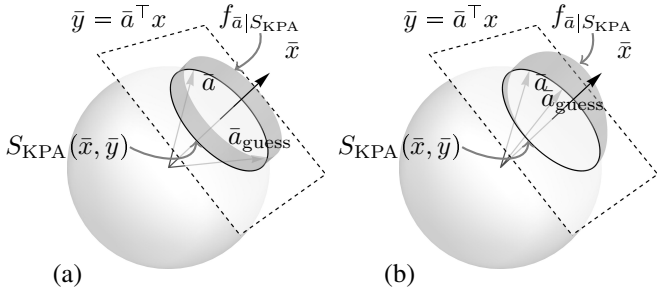

Fig. 4. The second effect of rakeness-based desing on KPA susceptibility: the distribution of the guessed sensing vectors is not uniform and typical guesses are closer.

radius $r = \sqrt{1 - (\bar{y}/\|\bar{x}\|)^2}$. We indicate such a sphere with $S_{\mathrm{KPA}}(\bar{x}, \bar{y})$. This structure helps visualizing the two effects that contribute to make rakeness-based CS more susceptible to KPA with respect to conventional CS.

First, to increase compression, rakeness-based design flow increases the average magnitude of the measurement, i.e., the average magnitude of the scalar product between $\bar{a}$ and $\bar{x}$. Since $\bar{x}$ is fixed and $\|\bar{a}\| = 1$ this means decreasing the angle of the cone defining $S_{\mathrm{KPA}}(\bar{x}, \bar{y})$ thus decreasing its radius. Candidates are therefore expected to be closer to each other. This effect is visualized in Figure 3-(b).

As a second effect note that, in classical CS, the normalized sensing vectors end up to be uniformly distributed over the n-dimensional unit sphere and thus on $S_{\mathrm{KPA}}(\bar{x}, \bar{y})$ so that all relative positions between $\bar{a}$ and $\bar{a}_{\mathrm{guess}}$ are equally probable.

On the contrary, rakeness-based design gives a non-isotropic distribution to the sensing vectors that, when restricted to $S_{\mathrm{KPA}}(\bar{x}, \bar{y})$, is also non-uniform. With this, it is more likely that $\bar{a}$ and $\bar{a}_{\mathrm{guess}}$ both appear close to the points in which the restricted PDF has a maximum, thus decreasing their average difference. This situation is visualized in Figure 4 where two profiles for the conditioned PDF $f_{\bar{a}|S_{\mathrm{KPA}}(\bar{x},\bar{y})}$ are compared.

Our path to consider both the effects will be made of an approximation step and some bounding steps, the idea being that we are mainly interested in the probability

$$P_{\mathrm{guess-Ok}}(h) = \Pr\left\{\frac{1}{n}\Delta(a, a_{\mathrm{guess}}) \leqslant h\right\} \tag{12}$$

from some small fraction of entries $h \ll 1$. To allow an analytical development, such a probability will be approximated and bounded by considering vectors that are not antipodal but Gaussian and with the same correlation matrix $\mathcal{A}$ as the antipodal rows.

We base our method on few mathematical facts that we state as Lemmas whose proofs are given in the Appendix. The results of the Lemmas are used in a less formalized path leading to the final approximation.

Lemma 1. *If γ is an n-dimensional, zero-mean Gaussian vector with correlation a matrix $\mathcal{A}$ equal to the one of an antipodal vector, and $\bar{\gamma} = \gamma/\sqrt{n}$ then*

$$\|\bar{\gamma}\|^2 \overset{n\to\infty}{\sim} \mathcal{N}\left(1, \frac{2\xi_a}{n}\right)$$

Lemma 1 suggests relaxing the antipodal a into γ since, as $n \to \infty$, $\bar{\gamma}$ has almost surely a unit length like $\bar{a}$ and the two vectors obey the same first and second-order statistics.

Yet, the attacker does not work with a generic $\bar{a}$ (and thus with a generic relaxed $\bar{\gamma}$) but starts from guessed row vectors a_{guess} generated with the same statistics that generates the true one a that also match the measurements equation ensuring that $a_{\mathrm{guess}}^\top x = y = a^\top x$, i.e., that $\bar{a}, \bar{a}_{\mathrm{guess}} \in S_{\mathrm{KPA}}(\bar{x}, \bar{y})$, that is an $n-1$-dimensional sphere centered (and orthogonal to) $c = \bar{y}\bar{x}/\|\bar{x}\|^2$.

From our relaxed point of view, we may model this by considering vectors $\gamma|\{y = \gamma^\top x\}$, i.e., vectors γ conditioned to the hyperplane passing through c and orthogonal to it. Actually, since the distance between any two such vectors does not change if we shift them along the direction orthogonal to the hyperplane, it is convenient to consider vectors $\tilde{\gamma} = \gamma|\{\bar{y} = \bar{\gamma}^\top \bar{x}\} - e$ with $e = \mathbf{E}[\gamma|\{\bar{y} = \bar{\gamma}^\top \bar{x}\}]$ to discard the average. Note that, in general $e \neq c$.

The following Lemma 2 bounds some first- and second-order statistics of vectors $\tilde{\gamma}$.

Lemma 2. *The random vectors $\tilde{\gamma} = \gamma|\{\bar{y} = \bar{\gamma}^\top \bar{x}\} - e$ are zero-mean Gaussian vectors with correlation matrix $\tilde{\mathcal{A}}$ whose eigenvalues $\tilde{\lambda}_j$ satisfy*

$$\frac{1}{n-1}\sum_{j=0}^{n-2} \tilde{\lambda}_j \overset{n\to\infty}{\geqslant} \tilde{\mu}$$
$$\frac{1}{n-1}\sum_{j=0}^{n-2} \tilde{\lambda}_j^2 \overset{n\to\infty}{\leqslant} \tilde{\sigma}^2$$

with

$$\tilde{\mu} = 1 - \frac{1 + (\xi_x - 1)\ell}{n-1} \tag{13}$$

$$\tilde{\sigma}^2 = 1 + (\xi_x - 1)\,\ell + \frac{\left[1 + (\xi_x - 1)\,\ell\right]^2}{n-1} \tag{14}$$

Starting from $\tilde{\gamma}$, we may consider the normalized $\tilde{\bar{\gamma}} = \tilde{\gamma}/\sqrt{n-1}$ as relaxations of $\bar{a}_{\mathrm{guess}}$ and $\bar{a}$ when $\bar{a}_{\mathrm{guess}}$ is produced by the attacker. This is intuitively represented in Figure 5.

Lemma 3. *Let's $\mathcal{G}$ be a $d \times d$ correlation matrix $\mathcal{G}$ of a d-dimensional random vector. Let the eigenvalues ζ_j of $\mathcal{G}$ be such that*

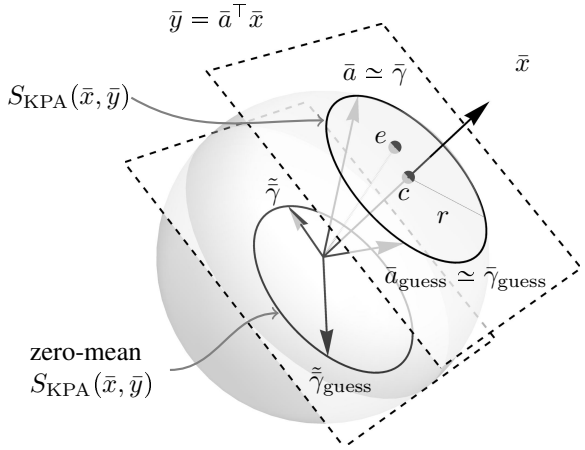

Fig. 5. Normalized true $\bar{a}$ and guessed $\bar{a}_{\mathrm{guess}}$ rows approximated by Gaussian vectors $\bar{\gamma}$ and $\bar{\gamma}_{\mathrm{guess}}$ and restricted to $S_{\mathrm{KPA}}(\bar{x}, \bar{y})$ to model the attack. Since we are interested in the difference between vectors we may translate $S_{\mathrm{KPA}}(\bar{x}, \bar{y})$ so that its vectors have zero mean in stead of e.

$$\begin{aligned} \lim_{d\to\infty} \frac{1}{d} \sum_{j=0}^{d-1} \zeta_j &= \mu \\ \lim_{d\to\infty} \frac{1}{d} \sum_{j=0}^{d-1} \zeta_j^2 &= \sigma^2 \\ \lim_{d\to\infty} \frac{1}{d^{3/2}} \sum_{j=0}^{d-1} \zeta_j^3 &= 0 \end{aligned}$$

for some quantities $0 < \mu, \sigma^2 < \infty$.

If g' *and* g'' *are independent and such that* $\sqrt{d}g', \sqrt{d}g'' \sim \mathcal{N}(0, \mathcal{G})$*, then*

$$\|g' - g''\|^2 \overset{d\to\infty}{\sim} \mathcal{N}\left(2\mu, \frac{8\sigma^2}{d}\right) \tag{15}$$

Convergence is such that the maximum difference between the cumulative distribution $\Phi_{\|g'-g''\|^2}$ *and the cumulative distribution* $\Phi_{2\mu, 8\sigma^2/d}$ *of the limit Gaussian is bounded by*

$$\left\| \Phi_{\|g'-g''\|^2} - \Phi_{2\mu, 8\sigma^2/d} \right\|_\infty \leqslant \frac{39}{d^{3/2}} \sum_{j=0}^{d-1} \zeta_j^3$$

Clearly, the vectors $\tilde{\gamma}$ satisfy the assumptions of Lemma 3 where the exact values of μ and σ^2 are not known but are bounded by Lemma 2. This allows to give an approximate bound on the probability that once the attacker generates a guess a_{guess} satisfying the measurements equation, its Hamming distance from the true a is small.

In fact, we may start from (11) and use the relaxation described above to say that the vectors a satisfying the measurement equation distribute approximately as the vectors $\tilde{\gamma}$.

The main issue in doing this, is that a_{guess} is a discrete-valued random vector while its approximation $\tilde{\gamma}$ is a continuous-valued one. For the two distributions to be alike, it is necessary that the number of possible a_{guess} is very large, i.e., that the measurement equation has a large number of possible solutions. In [3] a mathematical analogy between KPA and the well-known subset-sum problem is established to leverage what is known from the statistical approach to combinatorial problems. In particular, one gets that, if b_x is the number of bits with which one encodes the absolute value of the samples of the signal, the ratio $\delta = n/b_x$ (indicated as *density*) is the key parameter deciding how many rows potentially satisfy the measurement equation.

Roughly speaking, if δ is larger than few units, then the number of solutions tend to be extremely high and the approximation improves. Some empirical evidence on this phenomenon will be given in the Examples section. It is here enough to notice that our applicative framework implies that the samples of x come from a sensor and thus are typically encoded with $b_x \leqslant 20$ while n is easily in the order of the hundreds if not the thousands, yielding δ well in excess of 5, and most probably larger than 10.

As a consequence, we may apply Lemma 3 to say that $\frac{1}{n}\Delta(a, a_{\mathrm{guess}})$ is approximately distributed as $\mathcal{N}\left(\frac{\mu}{2}, \frac{\sigma^2}{2(n-1)}\right)$ and thus from (12)

$$P_{\mathrm{guess-Ok}}(h) \overset{n\to\infty}{\simeq} \frac{1}{2}\mathrm{erfc}\left(\frac{\frac{\mu}{2} - h}{\sqrt{\frac{\sigma^2}{n-1}}}\right) \tag{16}$$

Starting from such an estimation, we may exploit the fact that erfc is decreasing in its argument, as well as Lemma 2 to get

$$P_{\mathrm{guess-Ok}}(h) \overset{n\to\infty}{\leqslant} \frac{1}{2}\mathrm{erfc}\left(\frac{\frac{\tilde{\mu}}{2} - h}{\sqrt{\frac{\tilde{\sigma}^2}{n-1}}}\right) \tag{17}$$

for all those values h such that the argument of erfc in (17) is smaller than the argument of erfc in (16), i.e., for

$$h \leqslant \frac{\tilde{\mu}}{2} \frac{\frac{\mu}{\tilde{\mu}}\sqrt{\frac{\tilde{\sigma}^2}{\sigma^2}} - 1}{\sqrt{\frac{\tilde{\sigma}^2}{\sigma^2}} - 1}$$

Though it is not possible to compute such a limitation on h given that μ and σ^2 are unknown, it is safe to say that (17) holds for the small h that are of interest. Such an expression with the parameters in (13) and (14) can be used to give some security guarantee as it ensures that the probability of success of a KPA does not exceeed a certain level controlled by the features of the signal (ξ_x) and the level of adaptation we used in the design (ℓ).

VII. Examples

In this Section we apply rakeness-based CS to the acquisition of a number of signals and numerically assess the susceptibility of the resulting encoding to COAs and KPAs with the double aim of validating the theoretical predictions and quantifying the computational resources needed by the attacker to break the encryption.

We consider both synthetic and real-world signals. Relatively low-dimensional, synthetic signals are generated ad-hoc to be both sparse and localized and are used to explore the

design space and demonstrate the general trade-off between compression performance and security, as well as the capability of our theoretical model to give guarantees on achievable security. True ECG tracks are taken from [33], [34] to show that the whole machinery applies to real-world cases in which the dimensionality is large enough to validate asymptotic trends.

For synthetic signals $n = 64$, localization is obtained directly in the sparse domain considering $x = Ss$, with S being the $n \times n$ Discrete Cosine Transform matrix, and considering non-equal probabilities $\Pr\{s_j \neq 0\}$ that increase as the frequency of the corresponding columns of S increase.

The $\kappa = 6$ non-null components in s are randomly selected according to such probabilities. The magnitudes of non-null components are taken as independent random values uniformly distributed in $[-0.5, 0.5]$.

Clearly, directions associated to columns of S corresponding to larger probabilities are those along which x puts more energy. The resulting signal is therefore high-pass and its localization (1) is $\mathfrak{L}_x = 0.035$.

The samples in x are first quantized using $b_x = 12\,\text{bit}$ and then encoded in the measurement vector $y = Ax$. Since each row of A is made of $n = 64$ entries, each entry of y is encoded with $b_y = 12 + \log_2(n) = 18\,\text{bit}$. This is a conservative choice that does not take into account the possiblity to re-encode the entries of y to save bits (see, e.g., [14], [35]) that is out of the scope of this work.

Each row of the matrix A is generated as an independent antipodal vector with a correlation $\mathcal{A}$ obtained by a rakeness-based design flow parameterized by $\ell \in [0, 0.25]$. Recall that when $\ell = 0$ rakeness-based design yields the same sensing vectors as conventional CS. As ℓ increases, the statistic of the sensing vectors becomes more and more adapted to the signal to acquire. Hence, considering systems with different values of ℓ allow us to explore the trade-off between adaptation (and thus compression efficiency) and security.

As far as ECGs are concerned, we consider the record #100 of the MIT online database [34] included in the Physionet project [33]. Such a signal comes as a sequence of samples at a rate of 360 sample/s, each of them quantized by using 11 bit. From that stream we extract 1000 windows of $n = 256$ samples each. The localization of these signals is $\mathfrak{L}_x = 0.0242$.

The statistical characterization of this signal, necessary to generate the sensing rows a accordingly to the rakeness approach, has been taken from [13], from where we have also taken the suggested value $\ell = 0.25$.

A. Empirical evidence on COAs

According to (10), measurements asymptotically distribute as a zero-mean Gaussians with variance $\left(1 + \sqrt{\ell}\,(\xi_x - 1)\right) W_x$. Hence, any attack that relies on the statistical analysis of the ciphertex cannot extract any information but the variance of distribution. Since n and ℓ are known to the attacker, what actually leaks is only the

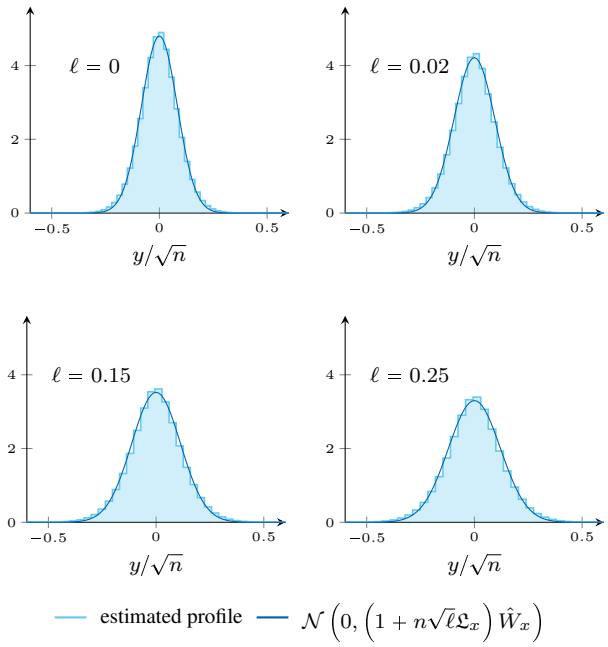

Fig. 6. Comparison between the actual probability distribution of the measurements taken from the synthetic signal and the asymptotic Gaussian trend predicted by the theory.

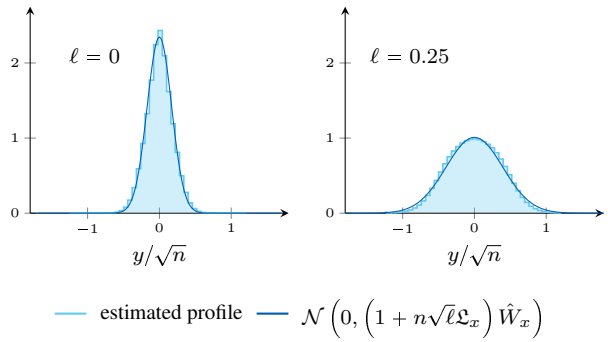

Fig. 7. Comparison between the actual probability distribution of the measurements taken from the real-world ECG tracks and the asymptotic Gaussian trend predicted by the theory.

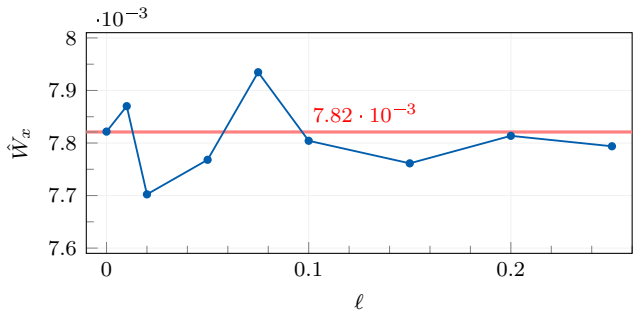

Fig. 8. Effectiveness of COA in estimating the average signal energy W_x that is the only leaking information.

estimation $\hat{W}_x$ of the actual signal power W_x computed as follows

$$\hat{W}_x = \frac{\hat{\sigma}_y^2}{1 + \sqrt{\ell}(\xi_x - 1)}$$

where the measurement variance $\hat{\sigma}_y^2$ is computed over different successive ciphertexts.

In Figure 6 we compare the empirical distribution of 10^8 sample measurements with the asymptotic distribution predicted by the theory for some values of ℓ. Figures 7 does the same comparison in the ECG case considering only the classical CS ($\ell = 0$) and rakeness-based CS with $\ell = 0.25$.

In Figure 8 we show how much the estimation of the average energy of the signal, obtained by dividing the variance of the measurements by $1+n\sqrt{l}\mathfrak{L}_x$, conforms to the true value $W_x = 7.81 \times 10^{-3}$ for different values of ℓ.

The same estimation can be tried for the ECG signals whose average energy is $W_x = 0.0324\,\mathrm{mV}^2$ giving the two very close values $\hat{W}_x = 0.0323\,\mathrm{mV}^2$ for classical CS and $\hat{W}_x = 0.0326\,\mathrm{mV}^2$ for rakeness-based CS.

B. Empirical evidence on KPAs

Though this was not modeled in the derivation of the theoretical guarantee, to practically test the effectiveness of a KPA we must define how the attacker produces guesses a_{guess} that simultaneously satisfy the given measurement equation $y = a_{\mathrm{guess}}^\top x$ and feature the given correlation matrix $\mathcal{A}$.

Though more efficient methods may exist, we tackle such a problem in the simplest way, i.e., by generating candidate vectors $a_{\mathrm{candidate}}$ using the same generator used by the encoder to produce the true row a, and promoting such candidates to proper guesses $a_{\mathrm{guess}} = a_{\mathrm{candidate}}$ when they satisfy the given measurement equation.

Figure 9 gives a quantitative assessment of the trade-off that can be expected between compression and security against KPAs. Empirical statistics are computed based on Monte Carlo simulations that generate a total of 1.5×10^{10} candidates.

Figure 9-(a) reports the Average Reconstruction Signal to Noise Ratio (ARSNR) that can be achieved applying a rakeness-based design with different values of ℓ. If x is the original signal, $y = Ax$ and $\hat{x}$ is the signal recovered by the decoder using y, A and the sparsity assumption, the ARSNR is estimated as the empirical mean of $\|x\|^2/\|\hat{x}-x\|^2$. Higher curves in Figure 9-(a) correspond to better performance as the quality of reconstruction is higher given the same number of measurements. This fact is usually exploited the other way around, i.e., by fixing a required ARSNR and using rakeness to decrease the number of measurement and, ultimately, the number of bits to transmit.

In our toy case, assuming we need an ARSNR of 60 dB, rakeness-based design helps reducing the number of measurements (see the Table at the bottom of Figure 9) from the 32 needed by classical CS ($\ell = 0$) to the 23 using $\ell = 0.25$. Though a thorough optimization of compression is out of the scope of this paper, it is clear that in the observed range, larger values of ℓ correspond to larger compressions.

Coming to KPAs, in Figure 9-(b) we plot the probability that a candidate row satisfies the measurement equation and has a Hamming distance from the true one not larger than a certain H. In this case, higher curves mean lower security since it is more probable that the attacker's guesses are good approximations of the true row. As highlighted in the zoom window, as ℓ increases such a probability increases.

To assess whether the theoretical guarantees are able to give a quantitative prediction of the level of security against KPAs, note that the probability of success of an attack is

$$P_{\mathrm{KPA-Ok}}(H) = P_{\mathrm{guess-Ok}}\left(\frac{H}{n}\right) \Pr\left\{y = a_{\mathrm{candidate}}^\top x\right\}$$

since we have a guess only when a candidate satisfies the measurement equation.

The probability $P_{\mathrm{guess-Ok}}$ is bounded by (17) while $\Pr\left\{y = a_{\mathrm{candidate}}^\top x\right\}$ depends on the specific method the attacker uses to generate guesses. Since probabilities are bounded by 1 we have

$$P_{\mathrm{KPA-Ok}}(H) \leqslant P_{\mathrm{guess-Ok}}\left(\frac{H}{n}\right)$$

and (17) can be used to guarantee a minimum level of security.

To match such a bound with the empirical evidence we collect, one needs to normalize plots like the one in Figure 9-(b) by $\Pr\left\{y = a_{\mathrm{candidate}}^\top x\right\}$ that can be trivially estimated. Figure 10 reports the resulting profiles for different values of ℓ, zooming in the low-H region that is the most interesting.

In each plot, the empirical probability is reported for values of H that have appeared at least once in the simulation. The empirical probability is then approximated by fitting (16) to the available data to provide a very good approximation that can be used to extrapolate the trend for lower values of H. The theoretical (17) is also reported showing that it is an upper bound for almost the whole left half of the H range.

Figure 11 reports the same profiles for the ECG case. Note how the upper bound is still valid for small H though the increase in n causes a strong decrease of the probability that guesses are close to the true row.

Beyond these examples, the general validity of (17) depends on an approximation that identifies the distribution of the antipodal vectors a satisfying the measurement equation $y = a^\top x$ with that of Gaussian vectors with the same second-order statistics. Section VI suggests that the quality of such an approximation depends on the density δ of the measurements equation, i.e. on the ratio between the dimensionality n of the vectors a and x, and the number of bits b_x used to encode the samples in x. Note that in the previous cases $\delta > 5$.

Though lower values of δ are unlikely to appear since in our framework b_x is the number of bits encoding a sample coming from the sensing of a real world signal, it is interesting to see what happens when δ is reduced. To do so, instead of increasing b_x to unfeasible values we scale down n keeping the statistic of the signals equal to that of the $n = 64$ case. In particular, we perform Monte Carlo simulations with $n = 24$ and $b_x = 4, 8, 12, 16$ to look into cases featuring a density decreasing from 6 to 1.

In Figure 12 we report the probability that a guess has a Hamming distances smaller than a certain H in the four

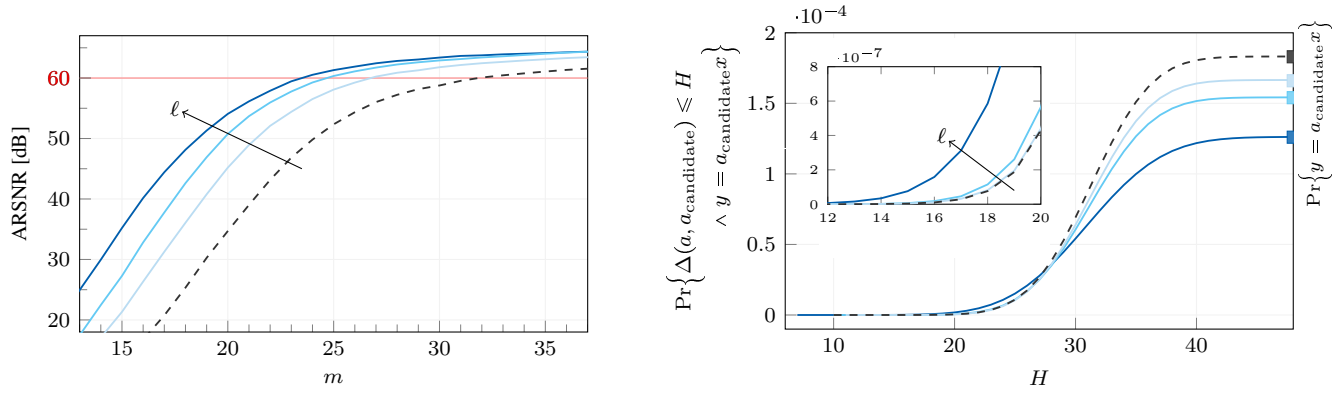

Fig. 9. Empirical evidence on the trade-off between compression and security against KPAs. Different curves correspond to different values of $\ell \in \{0, 0.01, 0.04, 0.25\}$ and thus of adaptation to the signal: (a) ARSNR plotted against the number of measurements m used for reconstruction; (b) Overall probability that a candidate row differs in not more than H positions from the true one. The table at the bottom reports the number of measurements needed to obtain ARSNR=60 dB and the probability that a candidate row satisfies the measurement equation.

different configurations, compared with their Gaussian fitting and with (17) that is the same in all cases. As b_x increases, the quality of the Gaussian fitting progressively degrades and the bound ceases to be valid. Yet, when b_x decreases (and thus the density increases) the Gaussian approximation becomes very good and the bound holds in the low-H region as expected.

Overall, in the $\delta > 5$ region, that is the most natural setting for a sensor, the bound is expected to work properly.

The availability of an analytical upper bound allows us to establish some guarantees on the computational security against KPAs.

In fact, if a KPA succeeds, the matrix A is revealed and we may assume that this allows to identify the key (i.e., the seed of the PRNG) and generate all the subsequent encoding matrices, thus exposing future plaintexts. It is sensible to assume that such an identification is possible when not more than H entries are erroneously guessed in each of the m rows of the matrix A, something that happens with a probability equal to $P^m_{\mathrm{KPA-Ok}}(H)$. A KPA may be tempted each time a ciphertext-plaintext pair is available to the attacker. The probability that none of T attacks succeeds is $\left(1 - P^m_{\mathrm{KPA-Ok}}(H)\right)^T$.

From a system level point of view, the most natural countermeasure to KPAs is changing the key. It is a potentially expensive countermeasure whose frequency is an indicator of the robustness of the encryption against the attacks. As a design guideline we may think that a key change is not needed until the probability of a successful KPA does not exceed a certain threshold $p_{\max}$.

By solving $1 - \left(1 - P^m_{\mathrm{KPA-Ok}}(H)\right)^T = p_{\max}$ for T we have that it is safe to keep the same key until the attacker has

$$T^* = \frac{\log\left(1 - p_{\max}\right)}{\log\left(1 - P^m_{\mathrm{KPA-Ok}}(H)\right)} \tag{18}$$

attack opportunities. Since (17) gives an upper bound on $P_{\mathrm{KPA-Ok}}(H)$ we are in the position of computing a lower bound for T^* that acts as a security guarantee.

	ℓ	m	$T^* \geqslant$ $H = 0$	$H = 1\%n$	$H = 10\%n$
synthetic n=64	0	32	3.7×10^{241}	8.1×10^{232}	2.6×10^{162}
	0.01	27	1.4×10^{199}	9.7×10^{191}	6.0×10^{133}
	0.04	25	5.1×10^{173}	3.0×10^{167}	6.4×10^{116}
	0.25	23	8.3×10^{114}	7.8×10^{110}	1.1×10^{78}
ECG n=256	0	134	1.8×10^{3857}	2.1×10^{3710}	7.4×10^{2519}
	0.25	74	3.2×10^{560}	3.9×10^{540}	4.3×10^{378}

TABLE I
THE NUMBER OF ATTACK OPPORTUNITIES THAT MAY BE SAFELY TOLERATED WITHOUT LEAVING TO THE ATTACKER A PROBABILITY OF SUCCESS GREATER THAN 10^{-4}. SUCCESS IS DEFINED AS THE IDENTIFICATION OF THE ROWS OF THE MATRIX A WITH A NUMBER OF ERRORS NOT LARGER THAN H.

Table I reports the number of attack opportunities that can be safely tolerated while being guaranteed that the probability that the attacker succeeds remains lower than $p_{\max} = 10^{-4}$ for $H = 0$ (all entries of A must be matched), for $H = 0.01n$ (up to 1% of the entries of A may be mistaken), and for $H = 0.1n$ (up to 10% of the entries of A may be mistaken). In the rows devoted to ECG we take from [13] the number of measurements needed to reconstruct an ECG both with a classical encoding and with a rakeness-based encoding.

Though a security degradation is easily appreciable when ℓ increases, even the smallest safety interval in Table I, $T^* \geqslant 1.1 \times 10^{78}$, is of astronomical proportion. Even if one attack could be performed each nanosecond, the resulting time would be of the order of 10^{61} years.

VIII. Conclusion

CS can be used as both a compression and an encryption stage in sensor nodes with a limited resource budget. A properly optimized CS stage is able to yield significant compression with a very limited computational complexity. Such

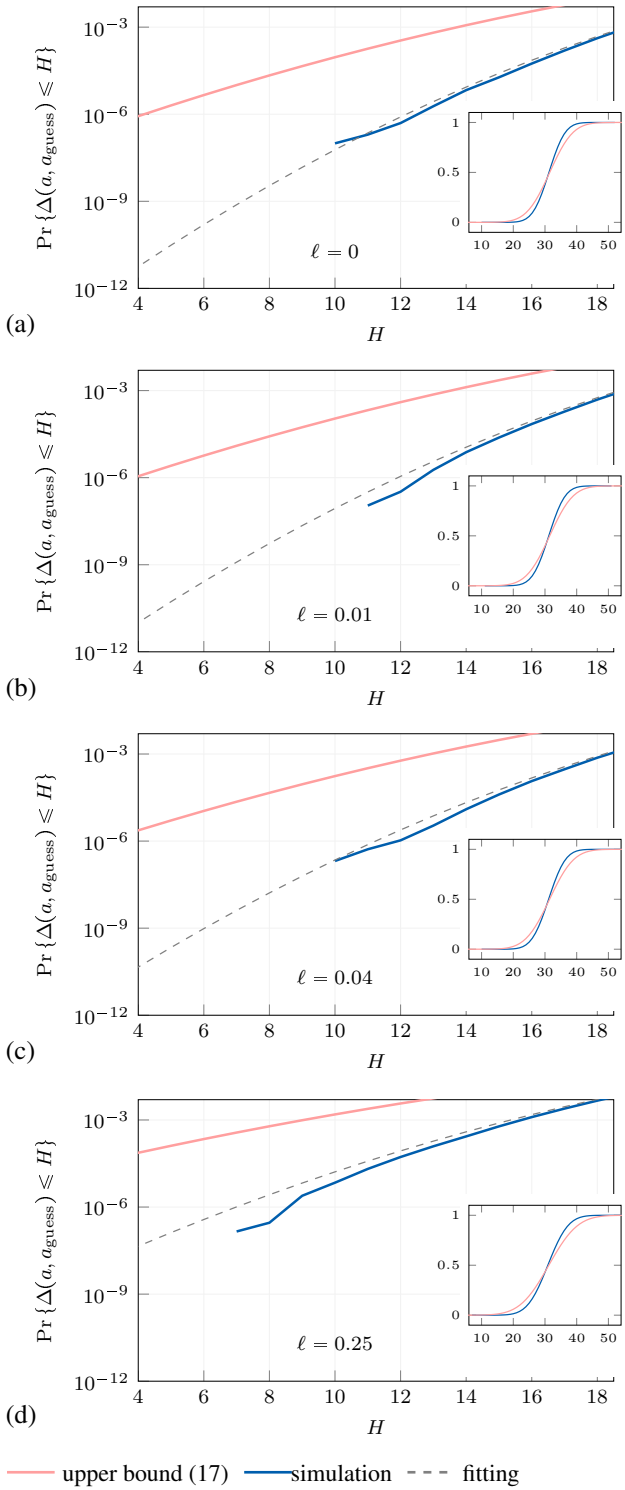

Fig. 10. Matching the numerical evidence collected in the synthetic case with the upper bound in (17).

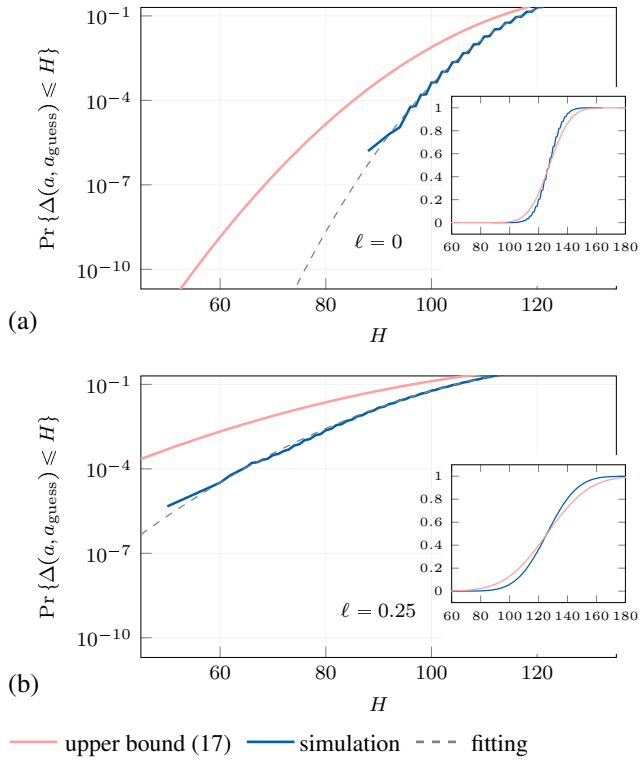

Fig. 11. Matching the numerical evidence collected in the ECG case with the upper bound in (17)

.

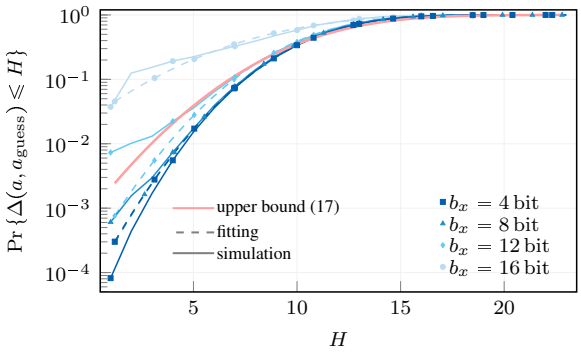

Fig. 12. Effect of density on the quality of the Gaussian approximation and, consequently, on the validity of (17).

an optimization relies on adapting the encoder to the statistical features of the signal to acquire. This partially compromises security since more information are available to the attacker. By means of some theoretical considerations, fully confirmed by the numerical evidence presented, we are nevertheless able to show that the loss in security does not prevent the method from exhibiting a noteworthy level of robustness with respect to classical attacks.

IX. Appendix

Proof of Lemma 1. Without any loss of generality we may think that the eigenvectors of $\mathcal{A}$ coincide with the coordinate axes so that the j-th component of γ is an independent Gaussian random variable with zero mean and variance λ_j. This allows to write

$$\|\bar{\gamma}\|^2 = \frac{1}{n}\|\gamma\|^2 = \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\left(\gamma_j^2 - \lambda_j\right) + \frac{1}{n}\sum_{j=0}^{n-1}\lambda_j$$

Each of the γ_j^2 is an independent, shifted χ^2 random variable with average λ_j and variance $2\lambda_j^2$, hence the inner part of the first term is a normalized sum of independent random variables with zero mean and variance $2\lambda_j^2$. The Central Limit Theorem implies that the limit behavior is that of a Gaussian with zero mean and variance

$$\frac{1}{n}\sum_{j=0}^{n-1}\mathbf{E}\left[\left(\gamma_j^2 - \lambda_j\right)^2\right] = \frac{1}{n}\sum_{j=0}^{n-1}\mathbf{E}[\gamma_j^4] - \lambda_j^2 = \frac{2}{n}\sum_{j=0}^{n-1}\lambda_j^2 = 2\xi_a^2$$

where we have exploited the fact that γ_j is Gaussian and thus $\mathbf{E}[\gamma_j^4] = 3\mathbf{E}^2[\gamma_j^2] = 3\lambda_j^2$. Since $\frac{1}{n}\sum_{j=0}^{n-1}\lambda_j = 1$ we have $\|\bar{\gamma}\| \overset{d\to\infty}{\sim} \mathcal{N}\left(1, \frac{2\xi_a}{n}\right)$ as required.

□

Proof of Lemma 2. Without any loss of generality we may think that c aligns with the first coordinate axis. Then exploit the symmetric and Töplitz structure of $\mathcal{A}$ to decompose it as

$$\mathcal{A} = \left(\begin{array}{c|c} 1 & \alpha^\top \\ \hline \alpha & \mathcal{A}_\lrcorner \end{array}\right)$$

The distribution of the vector γ conditioned to the knowledge that $\bar{y} = \bar{\gamma}^\top \bar{x}$ is still a Gaussian with mean $e = \left(\bar{y}/\|x\|^2\right)\alpha$ and covariance matrix [36, Section 3.4] $\mathcal{A}_\lrcorner - \alpha\alpha^\top$. Considering $\tilde{\gamma} = \gamma | S_{\mathrm{KPA}} - e$ leaves the covariance unaltered while making the average null. Hence $\tilde{\mathcal{A}} = \mathcal{A}_\lrcorner - \alpha\alpha^\top$ and

$$\tilde{\mathcal{A}}^2 = \mathcal{A}_\lrcorner^2 + \alpha\alpha^\top\alpha\alpha^\top - \mathcal{A}_\lrcorner\alpha\alpha^\top - \alpha\alpha^\top\mathcal{A}_\lrcorner$$

With this we have

$$\frac{1}{n-1}\sum_{j=0}^{n-2}\tilde{\lambda}_j = \frac{\mathrm{tr}\left(\tilde{\mathcal{A}}\right)}{n-1} = \frac{\mathrm{tr}\left(\mathcal{A}_\lrcorner\right) - \|\alpha\|^2}{n-1}$$

$$\frac{1}{n-1}\sum_{j=0}^{n-2}\tilde{\lambda}_j^2 = \frac{\mathrm{tr}\left(\tilde{\mathcal{A}}^2\right)}{n-1} = \frac{\mathrm{tr}\left(\mathcal{A}_\lrcorner{}^2\right) + \|\alpha\|^4 - 2\alpha^\top\mathcal{A}_\lrcorner\alpha}{n-1}$$

For the first term, note that $\mathrm{tr}\left(\mathcal{A}_\lrcorner\right) = n-1$ and that, if $\alpha = (\alpha_0, \ldots, \alpha_{n-2})^\top$ then the first row of $\mathcal{A}$ is $1, \alpha_0, \ldots, \alpha_{n-2}$. Hence,

$$\alpha_j = \mathcal{A}_{0,j+1} = \int_{-1/2}^{1/2} S_a(f) e^{2\pi \mathbf{i} f(j+1)} \mathrm{d}f$$

so that the Parseval equality gives

$$\|\alpha\|^2 = \sum_{j=0}^{n-2}\alpha_j^2 \overset{n\to\infty}{\approx} \int_{-1/2}^{1/2} S_a^2(f)\mathrm{d}f - 1 = \xi_a - 1 \qquad (19)$$

Therefore

$$\frac{1}{n-1}\sum_{j=0}^{n-2}\tilde{\lambda}_j \overset{n\to\infty}{\approx} \frac{n-1-\xi_a+1}{n-1} > 1 - \frac{\xi_a}{n-1}$$

For the second term, since $\tilde{\mathcal{A}}$ is positive semidefinite we have $\alpha^\top\tilde{\mathcal{A}}\alpha \geqslant 0$ and thus

$$\frac{1}{n-1}\sum_{j=0}^{n-2}\tilde{\lambda}_j^2 \leqslant \frac{\mathrm{tr}\left(\tilde{\mathcal{A}}^2\right) + \|\alpha\|^4}{n-1}$$

Yet, from the structure of $\tilde{\mathcal{A}}$ and (8) we know that

$$\mathrm{tr}\left(\tilde{\mathcal{A}}^2\right) \overset{n\to\infty}{\approx} (n-1)\xi_a \qquad (20)$$

.

and from (19) we have $\|\alpha\|^4 < \xi_a^2$ so that

$$\frac{1}{n-1}\sum_{j=0}^{n-2}\tilde{\lambda}_j^2 \overset{n\to\infty}{\leqslant} \xi_a + \frac{\xi_a^2}{n-1}$$

The thesis can be obtained recalling that $\xi_a = 1 + (\xi_x - 1)\ell$ (8). □

Proof of Lemma 3. Without any loss of generality we may think that the eigenvectors of $\mathcal{G}$ coincide with the coordinate axes so that the j-th component of $\sqrt{n}g'$ and $\sqrt{n}g''$ are independent Gaussian random variables with zero mean and variance ζ_j. This allows to write

$$\|g' - g''\|^2 = \frac{1}{\sqrt{d}}\frac{1}{\sqrt{d}}\sum_{j=0}^{d-1}\left[\left(\sqrt{d}g_j' - \sqrt{d}g_j''\right)^2 - 2\zeta_j\right] + \frac{2}{d}\sum_{j=0}^{d-1}\zeta_j$$

Since $\sqrt{d}g_j'$ and $\sqrt{d}g_j''$ are zero-mean independent Gaussian random variables with variance ζ_j, $\sqrt{d}g_j' - \sqrt{d}g_j''$ is a zero-mean Gaussian random variable with variance $2\zeta_j$.

Hence, the inner part of the first term can be written as

$$\frac{1}{\sqrt{d}}\sum_{j=0}^{d-1}2\zeta_j\left(\chi_j^2 - 1\right) \qquad (21)$$

where the χ_j^2 are independent χ-square random variable with one degree of freedom. Therefore, the above summands are zero-mean, with variance $8\zeta_j^2$ and 3rd-order moment $8\zeta_j^3\mathbf{E}\left[\left|\chi_j^2 - 1\right|^3\right]$. Straightforward calculations give

$$\mathbf{E}\left[\left|\chi_j^2 - 1\right|^3\right] = 8\left[1 + 3\sqrt{\frac{2}{e\pi}} - 2\mathrm{erf}\left(\frac{1}{\sqrt{2}}\right)\right] \simeq 8.69$$

so that the 3rd-order moments of the summands are $\simeq 69.53\zeta_j^3$.

Given the assumptions, the Central Limit Theorem ensures that (21) tends to a Gaussian random variable with zero mean and variance $8\sigma^2$. If $\Phi_{(21)}$ is the cumulative distribution function of the d-th term of the sequence in (21) and $\Phi_{0,8\sigma^2}$ is the cumulative distribution function of the limit Gaussian we know from [31] that

$$\left\|\Phi_{(21)}-\Phi_{0,8\sigma^2}\right\|_{\infty}\leqslant 0.56\times 69.54\frac{1}{d^{3/2}}\sum_{j=0}^{d-1}\zeta_j^3$$

Since $\|g'-g''\|^2$ can be obtained from (21) by scaling and offset, we know that $\|g'-g''\|^2 \overset{d\to\infty}{\sim} \mathcal{N}\left(2\mu, \frac{8\sigma^2}{d}\right)$ with the same convergence rate. □

REFERENCES

[1] S. Li, L.D. Xu and X. Wang, "Compressed Sensing Signal and Data Acquisition in Wireless Sensor Networks and Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2177–2186, 2013.

[2] E. Zimos, J.F.C. Mota, M.R.D. Rodrigues and N. Deligiannis, "Internet-of-Things data aggregation using compressed sensing with side information," *23rd International Conference on Telecommunications (ICT)*, pp. 1–5, 2016.

[3] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "On known-plaintext attacks to a compressed sensing-based encryption: A quantitative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2182–2195, 2015.

[4] T. Bianchi, V. Bioglio and E. Magli, "Analysis of One-Time Random Projections for Privacy Preserving Compressed Sensing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 313-327, 2016.

[5] M. Mangia, R. Rovatti, and G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 5, pp. 1001–1014, 2012.

[6] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "A rakeness-based design flow for analog-to-information conversion by compressive sensing," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2013, pp. 1360–1363.

[7] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[8] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[9] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[10] D. L. Donoho, A. Maleki and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, no. 45, pp. 18914–18919, 2009.

[11] L. Zheng, Z. Wu, M. Seok, X. Wang and Q. Liu, "High-Accuracy Compressed Sensing Decoder Based on Adaptive (ℓ_0,ℓ_1) Complex Approximate Message Passing: Cross-layer Design," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 10, pp. 1726-1736, Oct. 2016.

[12] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263.

[13] D. Bortolotti, M. Mangia, A. Bartolini, R. Rovatti, G. Setti and L. Benini, "An ultra-low power dual-mode ECG monitor for healthcare and wellness," *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1611-1616, 2015.

[14] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti and G. Setti, "A Case Study in Low-Complexity ECG Signal Encoding: How Compressing is Compressed Sensing?," in *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1743-1747, Oct. 2015.

[15] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, G. Setti, "Rakeness-Based Design of Low-Complexity Compressed Sensing," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, pp. 1201-1213, May 2017.

[16] F. Chen, F. Lim, O. Abari, A. Chandrakasan and V. Stojanovic, "Energy-Aware Design of Compressed Sensing Systems for Wireless Sensors Under Performance and Reliability Constraints," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 3, pp. 650-661, March 2013.

[17] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," in *2008 Forty Sixth Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2008, pp. 813–817, 2008.

[18] J. Barcelo-Llado, A. Morell, and G. Seco-Granados, "Amplify-and-Forward Compressed Sensing as a Physical-Layer Secrecy Solution in Wireless Sensor Networks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 839–850, 2014.

[19] A. Orsdemir, H. O. Altun, G. Sharma, and M. F. Bocko, "On the security and robustness of encryption via compressed sensing," in *2008 IEEE Military Communications Conference (MILCOM)*. IEEE, 2008, pp. 1–7.

[20] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Low-complexity multiclass encryption by compressed sensing," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2183–2195, 2014.

[21] Y. Zhang, L. Y. Zhang, J. Zhou, L. Liu, F. Chen and X. He, "A Review of Compressive Sensing in Information Security Field," in *IEEE Access*, vol. 4, no. , pp. 2507-2519, 2016.

[22] A. M. Abdulghani and E. Rodriguez-Villegas, "Compressive sensing: From "Compressing while Sampling" to "Compressing and Securing while Sampling"," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 1127-1130, 2010.

[23] R. Rovatti, G. Mazzini, G. Setti, "Memory-m Antipodal Processes: Spectral Analysis and Synthesis," IEEE Transactions on Circuits and Systems - I, vol. 56, n. 1, pp. 156-167, 2009.

[24] J. H. Van Vleck and D. Middleton, "The spectrum of clipped noise," *IEEE Proceedings*, vol. 54, no. 1, pp. 2–19, 1966.

[25] G. Jacovitti, A. Neri, G. Scarano, "Texture synthesis-by-analysis with hard-limited Gaussian processes," IEEE Transactions on Image Processing, vol.7, n.11, pp.1615–1621, Nov 1998.

[26] J. Xu, Y. Pi, Z. Cao, "Optimized projection matrix for compressive sensing," *EURASIP J. Adv. SignalProcess*, 2010.

[27] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, Chelsea Publishing Company, 1984

[28] A.N. Tikhomirov, "On the convergence rate in the Central Limit Theorem for weakly dependent random variables," Theory of probability and its applications, vol. XXV, n. 4, pp. 790–809, 1980

[29] A.C. Berry, "The Accuracy of the Gaussian Approximation to the Sum of Independent Variates," Transactions of the american Mathematical Society, vol. 49, n. 1, pp. 122-136, 1941

[30] V. Yu. Korolev, I.G. Shevtsova, "On the Upper Bound for the Absolute Constant in the Berry-Esseen Inequality," Theory of probability and its applications, vol. 54, n. 4, pp. 638–658, 2010

[31] I.G. Shevtsova, "An Improvement of Convergence Rate Estimates in the Lyapunov Theorem," Doklady Mathematics, vol. 82, n. 3, pp. 862–864, 2010 (*original russian text, I.G. Shevtsova, 2010, published in Doklady Akademii Nauk, 2010, Vol. 435, No. 1, pp. 26–28*)

[32] M. Mangia, F. Pareschi, R. Rovatti, G. Setti, "Security analysis of rakeness-based compressed sensing," IEEE International Symposium on Circuits and Systems 2016,

[33] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000

[34] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, May-June 2001.

[35] A. Moshtaghpour, L. Jacques, V. Cambareri, K. Degraux and C. De Vleeschouwer, "Consistent Basis Pursuit for Signal and Matrix Estimates in Quantized Compressed Sensing," in *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 25-29, Jan. 2016.

[36] M.L. Eaton, *Multivariate statistics: a vector space approach*, Institute of Mathematical Statistics, 2007

Mauro Mangia (S'09-M'13) received the B.Sc. and M.Sc. in Electronic Engineering and the Ph.D. degree in Information Technology from the University of Bologna (Bologna, Italy), respectively in 2005, 2009 and 2013. He is currently a Postdoctoral Researcher in the statistical signal processing group of ARCES - University of Bologna. In 2009 and 2012, he was a visiting Ph.D. student at the Ecole Polytechnique Federale de Lausanne (EPFL). His research interests are in nonlinear systems, compressed sensing, ultra-wideband systems, and systems biology. He was the recipient of the 2013 IEEE CAS Society Guillemin-Cauer Award and best student paper award at ISCAS2011. He is also the Web and Social Media Chair for ISCAS2018

Fabio Pareschi (S'05-M'08) received the Dr. Eng. degree (with honours) in Electronic Engineering from University of Ferrara, Italy, in 2001, and the Ph.D. in Information Technology under the European Doctorate Project (EDITH) from University of Bologna, Italy, in 2007. He is currently an Assistant Professor in the Department of Engineering, University of Ferrara. He is also a faculty member of ARCES - University of Bologna, Italy. He served as Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS - PART II (2010-2013). His research activity focuses on analog and mixed-mode electronic circuit design, statistical signal processing, random number generation and testing, and electromagnetic compatibility. He was recipient of the best paper award at ECCTD 2005 and the best student paper award at EMC Zurich 2005.

Riccardo Rovatti (M'99-SM'02-F'12) received the M.S. degree in Electronic Engineering and the Ph.D. degree in Electronics, Computer Science, and Telecommunications both from the University of Bologna, Italy in 1992 and 1996, respectively. He is now a Full Professor of Electronics at the University of Bologna. He is the author of approximately 300 technical contributions to international conferences and journals, and of two volumes. His research focuses on mathematical and applicative aspects of statistical signal processing and on the application of statistics to nonlinear dynamical systems. He received the 2004 IEEE CAS Society Darlington Award, the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as the best paper award at ECCTD 2005, and the best student paper award at EMC Zurich 2005 and ISCAS 2011. He was elected IEEE Fellow in 2012 for contributions to nonlinear and statistical signal processing applied to electronic systems.

Gianluca Setti (S'89-M'91-SM'02-F'06) received a Ph.D. degree in Electronic Engineering and Computer Science from the University of Bologna in 1997. Since 1997 he has been with the School of Engineering at the University of Ferrara, Italy, where he is currently a Professor of Circuit Theory and Analog Electronics and is also a permanent faculty member of ARCES, University of Bologna. His research interests include nonlinear circuits, implementation and application of chaotic circuits and systems, electromagnetic compatibility, statistical signal processing and biomedical circuits and systems. Dr. Setti received the 2013 IEEE CAS Society Meritorious Service Award and co-recipient of the 2004 IEEE CAS Society Darlington Award, of the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as of the best paper award at ECCTD2005, and the best student paper award at EMCZurich2005 and at ISCAS2011. He held several editorial positions and served, in particular, as the Editor-in-Chief for the IEEE Transactions on Circuits and Systems - Part II (2006-2007) and of the IEEE Transactions on Circuits and Systems - Part I (2008-2009). Dr. Setti was the Technical Program Co-Chair ISCAS2007, ISCAS2008, ICECS2012, BioCAS2013 as well as the General Co-Chair of NOLTA2006 and ISCAS2018. He was Distinguished Lecturer of the IEEE CAS Society (2004-2005 and 2014-2015), a member of its Board of Governors (2005-2008), and he served as the 2010 President of CASS. He held several other volunteer positions for the IEEE and in 2013-2014 he was the first non North-American Vice President of the IEEE for Publication Services and Products.